Application of Information Technology and Statistical Process Control in Pharmaceutical Quality Assurance & Compliance

by

Asvin Srinivasan

B.S. Electrical Engineering, University of Florida, Gainesville, 1997

Submitted to the MIT Sloan School of Management and the Department of Electrical Engineering & Computer Science in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Electrical Engineering & Computer Science

In conjunction with the Leaders for Global Operations Program at the Massachusetts Institute of Technology

June 2011

Signature of Author ____
May 6, 2011
Department of Electrical Engineering & Computer Science, MIT Sloan School of Management

Certified by _____
Roy E. Welsch, Thesis Supervisor
Professor of Statistics and Management Science and Engineering Systems

Certified by _____
Duane S. Boning, Thesis Supervisor
Professor and Associate Head of the Department of Electrical Engineering and Computer Science

Accepted by _____
Leslie A. Kolodziejski, Professor of Electrical Engineering and Computer Science
Chair of Department Committee on Graduate Students

Accepted by _____
Debbie Berechman, Executive Director of MBA Program
MIT Sloan School of Management

This page intentionally left blank.

Application of Information Technology and Statistical Process Control in Pharmaceutical Quality Assurance & Compliance

by

Asvin Srinivasan

Submitted to the MIT Sloan School of Management and the Department of Electrical Engineering & Computer Science on May 6, 2011 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Electrical Engineering & Computer Science

Abstract

Recently, the FDA issued new quality guidelines (Q10) encouraging drug manufacturers to improve their quality monitoring procedures. This renewed focus on quality and risk management has prompted Novartis to re-evaluate their systems and procedures to ensure compliance with the proposed guidelines. The company has chosen to respond by introducing more advanced statistical analysis of the data they share with regulatory bodies through the Annual Product Review (APR). However, procedural changes alone cannot bring about the needed innovation. Currently, too much time is spent on data consolidation and other non-value added tasks allowing less time for analysis. The solution is an Information Technology system with new procedures that will both improve process quality and increase productivity. The design proposed in this thesis utilizes statistical software that can analyze data securely, automatically generate graphs, and display alerts through an online dashboard. This *Decision Support System* will be integrated into Novartis's Global APR Automation project which aims to automate the generation of the entire APR document. A dashboard feature will allow processes to be monitored continuously instead of annually. The final version of the system will also include content management systems, business warehousing, audit validation and business intelligence tools. In addition to software, alternate statistical methods are proposed for evaluating critical processes that are either not in statistical control or lack normal distributions. These methods together with the new IT tools should help Novartis address process exceptions and reduce process variation without overloading the organization.

Thesis Supervisor: Roy E. Welsch
Title: Professor of Statistics and Management Science and Engineering Systems

Thesis Supervisor: Duane S. Boning
Title: Professor of Electrical Engineering and Computer Science

This page intentionally left blank.

Acknowledgments

I would like to thank my supervisor at Novartis Quality Assurance & Compliance, Eva Uebersax Keller, for her support and help with the research for this thesis. I would also like to thank the following individuals at Novartis for their support, Rafael Almoguera, Philippe Michaud, Arne Zilian, Stephanie Tozer, Zack Smith, Tatiana Yglesias, and Olivia Darmuzey.

This page intentionally left blank.

Table of Contents

List of Figures

1 Introduction

The pharmaceutical industry is known to be both capital intensive and highly regulated. However, despite the vast amounts of money spent on drug development, and stringent oversight by regulatory bodies, quality assurance monitoring is surprisingly primitive when compared with other industries. Batch quality failures for the pharmaceutical industry range from 5 to 15% [2] and waste has been reported to be as high as 50% [3]. In comparison, the semiconductor industry maintains waste well below 1% [4]. The table below provides a more detailed comparison between the two industries using a number of different quality metrics [1].

	Sigma	ppm Defects	Yield	Cost of Quality
Pharma	2σ	308,537	69.2%	25-35%
	3σ	66,807	93.3%	20-25%
	4σ	6,210	99.4%	12-18%
Semicon	5σ	233	99.98%	4-8%
	6σ	3.4	99.99966%	1-3%

Figure 1. Quality Comparison between Industries

It is clear that the pharmaceutical sector has a long way to go before it can achieve the type of quality control exhibited in the semiconductor industry. Even more troubling is the problem of out of specification batches failing to be detected by quality assurance and ending up in the hands of patients. A KPMG study reported that 72% of drug recalls are a result of manufacturing defects [5]. There are two reasons that are often cited for this lack of innovation in quality control tools.

1. High profit margins made increased investment in quality control tools unattractive.
2. Changes to manufacturing process require revalidation and extensive sampling which makes it uneconomical to continuously improve them.

These assumptions are beginning to change. As the industry goes through a period of slower growth and patent expirations, profit margins are beginning to decrease and quality costs will become increasingly significant. Meanwhile regulatory agencies are attempting to work with pharmaceutical firms to develop frameworks that allow more room for innovation.

1.1 Thesis Structure

Chapter 1 introduces the challenges facing the pharmaceutical industry in terms of quality control which provides the motivation for the thesis. The main objective of the thesis is described as well as a breakdown of the various research components. Chapter 2 gives a detailed background concerning the new regulatory guidelines issued by the FDA and their affect on the Annual Product Review document. The current Novartis procedure for statistical process control is presented along with its shortcomings. Chapter 3 provides a brief overview of some common statistical process control techniques as well as the key assumptions behind them. Chapter 4 is the data analysis section. It introduces the three quality parameters that have been collected for and describes the method for sampling and analyzing them. The data for each of these parameters is presented and underlying statistical problems are discussed. The chapter concludes with recommendations for how to best analyze these attributes and present them in a coherent and appropriate manner for the APR report. Chapter 5 is the IT section of the thesis. It describes the underlying enterprise architecture for the Quality Assurance group and the various systems that will make up the new Decision Support System. The chapter also describes the criteria for software evaluation and an explanation for the final software decision. Chapter 6 is a *Three Lens* analysis that highlights some of the organizational changes needed to fully utilize the new software tool. It highlights some of the cultural diversities between various groups at Novartis and how these may impact system implementation. Finally, Chapter 7 presents the conclusion of the thesis and next steps for the project.

1.2 Project Motivation

In an effort to spur pharmaceutical manufacturers to improve product quality and obtain better control over their manufacturing processes, the FDA in 2009 issued an industry guidance titled *Q10 Pharmaceutical Quality System.* The guidance describes a "pharmaceutical quality system to enhance the quality and availability of medicines" [7]. One of its key objectives is to encourage manufacturers to "develop and use effective monitoring and control systems for process performance and product quality, thereby providing assurance of continued suitability and capability of processes" [7]. The guidance impacts a number of areas including regulatory filings drug manufacturers must produce each year such as the Annual Product Review (required for compliance in the U.S) and the Product Quality Review (required for compliance in the E.U).

Every manufactured drug must have these reviews submitted periodically in order to meet regulatory compliance (and continue producing the drug). The Q10 guidelines affect these reviews because it suggests[1] manufacturers should now provide statistical analysis of parameters critical to the production process. The motivation behind this is to help manufacturers identify problems early, identify their root cause and take actions to prevent the problems from occurring in the future. The current system in place for performing these tasks can be improved so that it is better suited to meet these requirements. For example, the overall process for generating the APR report lacks an audit trail which can create data security issues. The structure of the databases also leads to large amounts of time to be spent on data consolidation instead of data analysis. Therefore a new system is required to correct these issues, and improve productivity.

1.3 Objective

This thesis proposes a framework for a Decision Support System (DSS) comprised of a backbone IT infrastrucutre and business intelligence tools that will help meet the quality system guidelines proposed above. The tool will automatically generate charts, tables, graphs for particular sections of the APR/PQR report, provide continuous monitoring of critical process parameters, alert users to outliers and trends which lead to deviations (using a dashboard) and provide a basis for assessing process capability when possible. Specific analytical methods are proposed for the three critical process parameters studied as part of the research.

1.4 Novartis Pharmaceuticals

The recommendations proposed in this thesis have been specifically designed for Novartis Pharmaceuticals, a division of Novartis A.G based in Basel, Switzerland. The research described in this thesis is the result of a six month internship with Quality Assurance & Compliance division within the Technical Operations Group. Research and data was gathered primarily from the Novartis manufacturing plant located in Stein, Switzerland. The Stein facility is the largest production plant within Novartis Pharmaceuticals. It is also a "Center of Excellence" which is a designation Novartis grants to plants that specialize is a certain area. Stein specializes in

[1] FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidance describes the Agency's current thinking on a topic and should be viewed only as a recommendation.

launching new drug products [5]. Almost all new Novartis drugs are first manufactured on a commercial scale at the Stein plant. During this initial campaign, the manufacturing process is fine tuned and any bottlenecks in the developmental stage are addressed to ensure high efficiencies and quality. Stein's role as a launch site for new drugs, means it is both a high throughput plant for blockbuster drugs, and a low volume, high mix plant (LVHM) for newer drugs. This complicates the design of automated tools because the high mix of drugs means the software tool must be able to handle many different data types. Shorter campaigns for new drugs also yield fewer data points which create issues when assessing process capability and stability. Therefore the quality assurance monitoring system must be robust enough to handle different production environments if it is to be successfully deployed.

1.5 Global APR/PQR Automation Project Integration

In June 2010, Novartis Pharmaceuticals approved a project to build a software tool to automate the generation of the entire APR/PQR report. The APR/PQR report consists of several chapters relating to various aspects of the drug manufacturing process (see Chapter 2 for detailed explanation). For the APR/PQR automation project to be a success, the tool must be able to automatically generate chart and graphs along with related capability indices. Two individuals head up the automation project team. One is responsible for representing the business needs while the other is responsible for IT project management. Both these individuals periodically report back to a steering committee that approves incremental funding and major project decisions such as selection of software vendors and changes to the project scope. Since statistical analysis is required for both the APR/PQR report and meeting the objectives of the Q10 Quality System guidelines, it has been proposed that the statistical platform and related online dashboard become part of the approved automation project. The steering committee recently approved the proposal. The decision means that the proposed statistical platform will eventually be deployed globally to all pharmaceutical manufacturing sites.

2 Background

This chapter begins with a background on the APR/PQR document and its objectives in terms of regulatory compliance. Aspects of the APR/PQR that are relevant to the thesis are highlighted. The chapter then describes current procedures for quality control and Good Manufacturing Practice (GMP) compliance at Novartis. The final section describes in detail the need for an updated decision support system.

2.1 Regulatory Compliance and the APR/PQR

Unlike the production of most other products such as microchips or clothes, pharmaceutical drugs are highly regulated by the governments of the countries in which they are sold due to their sensitive nature. In the United States, pharmaceutical drug manufacturing is governed by the Good Manufacturing Practice (GMP) guidelines which are enforced by the FDA, under Section 501(B) of the 1938 Food, Drug, and Cosmetic Act. GMP guidelines must be adhered to strictly and regulatory agencies reserve the right to inspect any drug production site they chose without prior notice in order to ensure compliance. As part of these guidelines, the FDA requires an annual review known as Product Annual Review (PAR) also referred to as the Annual Product Review (APR). The purpose of this proposed GMP requirement is to provide a mechanism for both the regulatory agency and drug manufacturer to review the quality standards for each drug product manufactured. In 2001, the FDA adopted and published the guidance for industry *Q7A Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients* [1]. While the guidance was formally issued by the FDA, it was developed within the "Expert Working Group (Quality) of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)." The guidance has been endorsed by the ICH Steering Committee and adopted by the regulatory bodies of the European Union, Japan, and the United States [7].

In the European Union, the EMEA is the regulatory body responsible for coordinating the scientific evaluation of the safety, efficacy, and quality of medicinal products. The EU GMP Guide is the document that provides the details supporting the principles of GMPs within the EU. In 2004, EMEA made available a draft revision of the EU GMP Guide proposing, for the first time, a requirement for a Product Quality Review. The Product Quality Review (PQR) has the

same objectives as the APR and its structure is similar to the APR. A comparison of objectives between the two regulatory reports is shown in Figure 2.

Table I: Review objectives.

Objectives	FDA: Product Annual Review 21 *CFR* 211.180(e)*	EC: Product Quality Review (1.5)*	ICH Q7A: Product Quality Review (2.5)*
Determine appropriateness of, and/or need to change, product specifications	Required	Required	Not specified
Appropriateness of starting material specifications	Not specified	Required	Not specified
Determine the need to change manufacturing procedures	Required	Not specified	Not specified
Determine the need to change manufacturing control procedures	Required	Not specified	Not specified
Verify consistency of the existing process	Not specified	Required	Required
Determine the need to revalidate the production process	Not specified	Required (also specified in EU GMP Annex 15)	Required (also specified in section 12.6)
Highlight trends	Expected, not specified	Required	Not specified
Identify product and process improvements	Not specified	Required	Not specified
Identify corrective actions	Expected, not specified	Required	Required

**CFR* denotes *Code of Federal Regulations*, EC denotes European Commission; and ICH denotes International Conference on Harmonization.

Figure 2. Comparison of Objectives between APR and PQR [8]

In order to ease the regulatory burdens, many manufacturers adopt the APR/PQR report which serves to meet the requirements of both reports. In essence, the APR/PQR is an annual retrospective revalidation of the manufacturing process. The procedures for performing a typical APR/PQR involve the review, analysis, and trending of historical data (i.e., data generated in the past 12 months), which fit the definition of retrospective process validation as defined in the FDA's validation guideline and the EU GMP Guide Annex 15 on qualification and validation (5, 6). Because the process of consolidating data and performing statistical analysis is both

computationally intensive and repetitive, Novartis has proposed that the process can be highly automated through the development of new processes and better integrated software.

2.1.1 Process Validation

When a new drug product is introduced in a manufacturing plant and scaled up for commercial sale, the production process must be validated to ensure it meets the requirements stated in the initial regulatory filing. The initial validation is based on three trial run batches and once these batches are shown to be within the specification, the process is validated and essentially locked in that form. Subsequent process changes are difficult. Anytime a change is requested to make a process more efficient, a manufacturer must file data and documents verifying that the new process does not impact product quality [9]. The APR/PQR must document all process changes and show the effect (if any) these have on the quality of the drug.

2.1.2 CAPA

In addition to aiding with process validation, the APR/PQR contains sections that can help identify trends and assist with Corrective and Preventative Actions (CAPA). CAPA is a concept within GMP that focuses on indentifying deviations and failures, investigating their "root cause" and taking preventative actions to ensure these sources of variation are eliminated in future batches.

2.2 Quality System Guidance Objectives

The purpose behind this project is to support Novartis so they can meet the guidelines set forth by the FDA in the Q10 Quality Systems guidance. While the Q10 guidelines contain a number of goals, the project focuses on targeting the following specific Q10 objectives [7]:

- Participate in the design, implementation and monitoring of the pharmaceutical quality system.

- Demonstrate strong and visible support for the pharmaceutical quality system and ensure its implementation throughout their organization.
- Key performance indicators should be identified and used to monitor the effectiveness of processes within the pharmaceutical quality system as described in Section 4 of the Q10 Guidelines.

The system developed as part of this project will allow production and Quality Assurance to meet these objectives in a time frame much faster than currently possible. The productivity gains will allow staff to spend more time investigating deviations and improving the process instead of gathering and analyzing data.

2.3 Current Quality Assurance Procedures

The current procedures for quality monitoring at Stein can be broken down into two categories. The first category governs the sampling, testing and data entry procedures that are part of the QA/QC technician's daily report. The second category governs the procedure for generating the APR/PQR reports. APR/PQR reports are compiled by the QA/Compliance department according to the quality directive written by the Global Quality Assurance department which must comply with GMP guidelines. The procedure for the trending of the predefined quality attributes was written by the Novartis IQP team and is described in Section 3.1, Chapter 2.

2.3.1 Process Map for Product Sampling and Testing

Finished product is randomly sampled from a production batch for a particular drug and delivered to the Quality Control lab[2]. Depending on whether the drug is packaged in a solid or liquid form, it is directed to a particular QA/QC lab. Lab technicians perform a variety of tests not only on the finished product, but also on the incoming raw material and active ingredients. The types of tests and the procedures for these tests will be described in Chapter 4. The data from these tests are automatically uploaded into a software package called Chromeleon. Any

[2] Batches are sampled to ensure independence.

data points outside the internal specification limits for the drug (referred to as out of expectation or OOE) are highlighted. Samples are re-tested depending on the procedures of the particular test. Chromeleon also computes a number of other relevant values for the sample batch including the min, max, mean and standard deviation values. Lab technicians then manually enter them into the SAP Quality Management (QM) database. Individual values are not copied to SAP at the moment.

2.3.2 IQP and the Current Procedure for APR/PQR

The Innovation Quality Productivity (IQP) group is an internal Novartis team that works with manufacturing, quality assurance and other stakeholders to introduce lean manufacturing and six-sigma practices. While the guidelines governing the overall APR/PQR report are documented in the global Novartis *Pharma Directive*, the section on Analytical Results is guided by procedures issues by IQP. The IQP team wrote the Standard Operating Procedure (SOP) for performing Process Capability Analysis using Minitab® Statistical Software [12]. The document "details the procedure to follow for the calculation and the statistical analysis of the process capability, using the Minitab® Statistical Software." Process capability is one of the main statistics Novartis lists in the Analytical Results section to indicate the performance of a particular process. The procedure for calculating process capability according to the SOP is summarized below.

1) Lab technicians must create a MS Excel spreadsheet with all testing monograph data compiled from various databases including SAP QM.

2) Minitab Statistical Software is then opened, and the excel spreadsheet data is copied and pasted into a Minitab worksheet (check for data integrity).

3) The Capability Six Pack Analysis (located under Quality Tools) is used to create a run chart, distribution chart, and the process capability statistics for a particular data set.

4) The excel spreadsheet and the Minitab file is then e-mailed to QA/CPL personnel that will be completing the Analytical Results section of the APR/PQR. Discussions with production and QA/QC are necessary for the interpretation of the data.

5) The graphs are copied and pasted into a MS Office Word document and then e-mailed to other parties who will work on subsequent sections of the document.

There are several major flaws with the current procedure.

- Consolidating data from various databases into an excel spreadsheet is time consuming (often taking more than three hours). It is non-value added since it simply involves replicating data already gathered. This reduces the amount of time personnel can spend on root cause investigations.

- All the data transfers between SAP QM, Excel, and Minitab are unsecure. This means data can accidentally be modified with no method of verifying data integrity. Data integrity is a required component of compliance. For example, if regulatory audits show discrepancies between data submitted on the APR/PQR and actual data collected, it could lead to major penalties.

- There is no secure audit trail for the APR/PQR document as it is assembled. This makes it difficult to track changes.

- The procedure does not provide much information on how to handle data sets that do not yield themselves to process capability assessment. Process Capability analysis requires data to be normally distributed and statistically stable. Yet there are many instances where one or neither of these requirements is possible. Alternate methods must be suggested for handling these data sets.

3 Statistical Process Control Basics

In this chapter, we will analyze statistical process control techniques as well as the current Novartis procedure for statistical charts and process performance. We will also analyze the underlying assumptions for control charts and process capability. This will have significant implications of the data sets discussed in Chapter 4.

3.1 Statistical Process Control

The use of Statistical Process Control in manufacturing goes back to the 1920s when William Shewhart (then at Bell Labs) first pioneered the use of control charts and other mathematical methods to improve the quality of telephones manufactured at the time [13]. Since then, numerous experts in both industry and academia have contributed to the field through more advanced techniques and methodologies. At its core, SPC consists of a set of techniques that quantify the variation within a manufacturing process and help users determine whether the variation is natural to the process or requires further investigation. The physical benefit of such a system is that it can help predict manufacturing problems before they physically manifest themselves as defects. The Novartis IQP group has adopted several SPC principles into the procedure for statistical analysis in the APR/PQR.

3.2 Chart Types

There are many different types of charts that can be used to analyze process data. A brief overview of common chart types is provided below.

3.2.1 Acceptance Chart

The simplest of all statistical charts, the acceptance chart simply graphs the sample data, provides the sample mean, and the customer specification limits for the process. The chart provides a visual method for ensuring data is within specification and not out of compliance. While the chart provides no information about process variation (and capability) it can help identify trends and potential problems in a qualitative sense.

3.2.2 X Chart and Process Control

The $\bar{X}$ chart also known as a $\bar{X}$ control chart is a graphical display for a quality characteristic that has been measured. If the data points represent sample averages, then $\bar{X}$ is

used to monitor the process mean. The chart contains a center line and an upper control limit line (UCL) as well as a lower control limit line (LCL). Typically, a process is assumed to be in-control, if all the sample data falls between the UCL and LCL. The UCL and LCL are typically calculated using the center line and the sample standard deviation. Typically, μ_w (the mean of the sample statistic w) is used for the center line and σ_w (the standard deviation of the sample statistic) is used to calculate the UCL and LCL. If a process is deemed to be within six-sigma, then all the sample values must fall between $+3\sigma_w$ and $-3\sigma_w$. More generally

$$Upper\ Control\ Line = \mu w + L\sigma w$$

$$Center\ Line = \mu w$$

$$Lower\ Comtrol\ Line = \mu w - L\sigma w$$

where L is the distance from the control limits to the center line.

In Minitab, data points that fall outside these control limits are highlighted in red, indicating they are outliers. Outliers can also exist inside the control limits depending on the user defined rules. Novartis IQP has chosen four tests to determine if a data point is in control. These four rules are defined in Minitab and given below:

1. One point is more than three standard deviations from center line (outside of control limits).

2. Nine points in a row on the same side of the center line.

3. Six points in a row all increasing or decreasing.

4. Fourteen points in a row all alternating .

Rules 2-4 help identify non-random patterns that may be caused by underlying process issues. For example, nine points in a row on the same side may indicate that the process is beginning to drift in a particular direction. Action must be taken to determine the root cause.

3.3 Process Capability

Process capability refers to the uniformity of the process and is an integral part of six sigma analysis. It allows the user to measure the variability in a process using both the mean and the standard deviation of the sample data. Process capability can be assessed with or without regard to the process specifications. Major uses for the data gathered through process capability analysis include the following [13]:

- Predicting how well the process will hold the tolerances.
- Assisting process experts in modifying the process.
- Assisting in establishing the interval between sampling for process monitoring.
- Specifying the performance requirements for new equipment.
- Reducing the variability in a manufacturing process.

Process capability ratios provide a quantitative way to express process performance.

$$\widehat{C_{pu}} = \frac{USL - \hat{\mu}}{3\sigma}$$

$$\widehat{C_{pl}} = \frac{\hat{\mu} - LSL}{3\sigma}$$

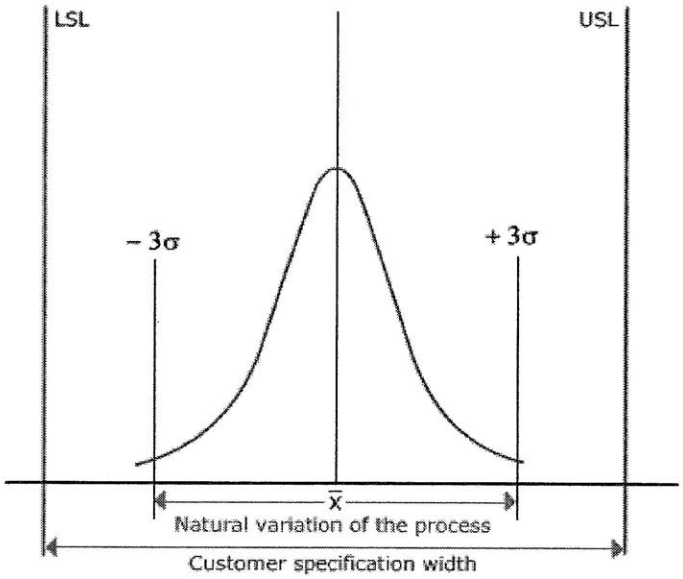

Figure 3. Histogram with Control Limits

The ratio C_p does not account for whether the mean of the sample data is centered between the control limits. It only measures the spread of the specifications relative to the six-sigma spread of the process. To account for the centering of a process, the term C_{pk} is defined as the minimum of C_{pu} and C_{pl}. The difference in magnitude between C_p and C_{pk} indicates how well the process is centered. If the two terms are equal then the process is perfectly centered between the upper and lower specification limit. The term C_p is therefore referred to as process capability and the term C_{pk} is referred to as actual capability.

3.3.1 Independence

In order to use control charts effectively, it is assumed that data sets are independently distributed. If the quality characteristics represent even low levels of positive correlation, control charts will give misleading results in the form of too many false alarms [13]. When we consider the Novartis manufacturing processes we must be careful to ensure that independence of observations is maintained. Sampling procedures currently in place help ensure that the samples

are representative of the batch. Autocorrelation of the data over an annual period shows fairly low autocorrelation, indicating that the process appears to be stochastic.

3.3.2 Assumptions for Process Capability

There are two critical assumptions for process capability calculations. One is the assumption that the underlying process produces a normal distribution of values. The other assumption is that the process is in statistical process control. Testing for normal distribution can be performed using a normal probability plot. If the plot fails to yield a nearly straight line then we can reject the hypothesis (with some confidence level) that the underlying data is normal. For a process to be in-control, the data points must fall between the three sigma limits, with a corresponding expected false alarm rate. Furthermore, process capability is attempting to measure variance due to natural causes so all excursions with known root causes (such as a non-calibrated machine) must be excluded from the calculation. If either of these assumptions is violated, then the estimates for process capability ratios (based on sample data) may not accurately reflect the underlying process. This is a key issue for the data sets analyzed in this thesis, and in Chapter 4 we will investigate methods for handling data sets that violate these assumptions.

4 Data Analysis

Quality Assurance/ Quality Control is responsible for performing a number of tests on drug products. However, the APR/PQR only requires manufacturers to provide process data for metrics deemed as "critical parameters". It is therefore up to each firm to decide which parameters are critical for a particular product. For the purposes of this project, it was important to narrow down the field to a few parameters and determine the best way to analyze these data sets using new methodologies. The parameters chosen for this task are dissolution, assay and content uniformity. These are critical parameters for products produced by the solids department. These parameters will be described in the sections below.

4.1 Dissolution

Dissolution refers to the rate at which a tablet or capsule dissolves. This is an important parameter for solids because the rate of dissolution reflects the rate of the active ingredient which is available for later absorption in the blood stream. Every solid drug product (typically a tablet or capsule) has a Q value which refers to the amount of active ingredient that must be dissolved within a specified time. The U.S Pharmacopeia has strict guidelines for performing the dissolution test [14]. The test involves placing a sample in a lab apparatus with a solvent that simulates the dissolution of active ingredients in the blood stream. Careful attention must be paid to the pH of the solvent, the temperature and other factors to ensure that the tests are accurate. There are also separate methods for immediate release dosage and extended release dosage forms. An acceptance table (shown below) explains how samples are tested and measured. If the samples fail stage 1, there are two additional opportunities for the batch to pass. Essentially each stage takes a wider set of samples and relaxes the threshold for compliance.

Stage	Number Tested	Acceptance Criteria
S_1	6	Each unit is not less than Q + 5%
S_2	6	Average of 12 units ($S_1 + S_2$) is equal to or greater than Q, and no unit is less than Q – 15%
S_3	12	Average of 24 units ($S_1 + S_2 + S_3$) is equal to or greater than Q, not more than 2 units are less than Q – 15%, and no unit is less than Q – 25%.

Figure 4. Acceptance Table for Dissolution [15]

To perform the dissolution test, six samples are initially chosen and tested. After a specified time, the percentage of dissolved active ingredient is measured for each sample, and these percentages must be no less than 5% of the stated Q value. If any samples fail this test, there are similar requirements for stage 2 and stage 3 with additional samples. Ideally, we would like all the individual values to be stored in the database for use in the statistical analysis. Unfortunately, in the current set-up, only the minima values are stored.

4.2 Assay

The assay test is performed to determine the amount of active ingredient in a tablet or capsule and ensure that it is in compliance with the drug label [17]. The measure is typically reported as a percentage. The assay can also be thought of as the dosage strength or potency of a drug. The quantity that refers to 100% is set to a finite value so it is possible to have values greater than 100. Determining what the appropriate compliance limits for assay should be is difficult because of drug stability issues. For example suppose a label states that the amount of active ingredient in a tablet should be 95% of the stated active ingredient weight. When a batch of these tablets is initially produced, a sample might show it is 100%, which is well within the 95% compliance. However, it is known that many active ingredients degrade over time through a variety of processes including hydrolysis and oxidation [18]. Therefore, after a year on the shelf, this same tablet might only have 90% of the stated active ingredient and could be out of specification with the label. Incidentally, this is why most drugs have an expiry date. Drug stability is a major issue for manufacturers because it is difficult to simulate several years of

shelf life in the lab. Pharmaceutical firms must continue to sample batches many years after they are produced and if the amount of active ingredient falls below a threshold, the batch must be recalled. Therefore the internal release limits for assay are based on the corresponding stability tests to ensure active ingredient content remains within the customer specifications for the shelf life of the drug.

4.3 Content Uniformity

Content uniformity (also referred to as Uniformity of Dosage Unit) is defined as *the degree of uniformity in the amount of drug substance among dosage units* [16]. The test is performed to ensure the consistency of dosage units (for example tablets) and ensure that the drug substance content for each tablet in a batch is within a narrow range around the label content. In essence, it is a statistical test performed on the assays to ensure that there are no significant deviations among the samples in a batch. The U.S Pharmacopeia describes in detail the statistical calculations for determining the *acceptance value* for content uniformity based on the assay values for individual batch samples [16].

Acceptance Value (AV) is defined as

$$\mathrm{AV} = |M - \bar{X}| + ks$$

The variables in the acceptance value calculation can be summarized as follows.

In this equation, the $\bar{X}$ variable is the mean of the individual sample assay values expressed as a percentage of the label claim. M is a reference value that can either take on the value $\bar{X}$, 98.5%, or an arbitrary target value for assay depending on the range for $\bar{X}$. Typically the target value for the assay is 100% but this can change depending on other factors (such as drug stability). The variable k is an acceptability constant that is either 2.4 or 2.0 depending on the sample size. Finally s is the sample standard deviation.

$$\mathrm{s} = \sqrt{\frac{\sum_{i=1}^{n}[(\mathrm{x_i} - \bar{\mathrm{X}})^2]}{\mathrm{n} - 1}}$$

The term L1 is defined as the maximum allowed acceptance value. Unless otherwise specified, U.S Pharmacopeia defines L1 to be equal to 15.0. The typical sample size for the uniformity of dosage test is 10.

4.4 Data Analysis Issues

As stated in Chapter 2, the main goal of both the internal Novartis process capability guidelines and statistical process control is to be able to assess to the performance of a process and detect problems before they manifest themselves. Using Minitab and the Novartis Process Capability SOP, process capability for a number of attributes have been calculated and are shown below. Recall that the two requirements for calculating process capability are that the process must be statistically stable, and the data set under analysis must be normally distributed.

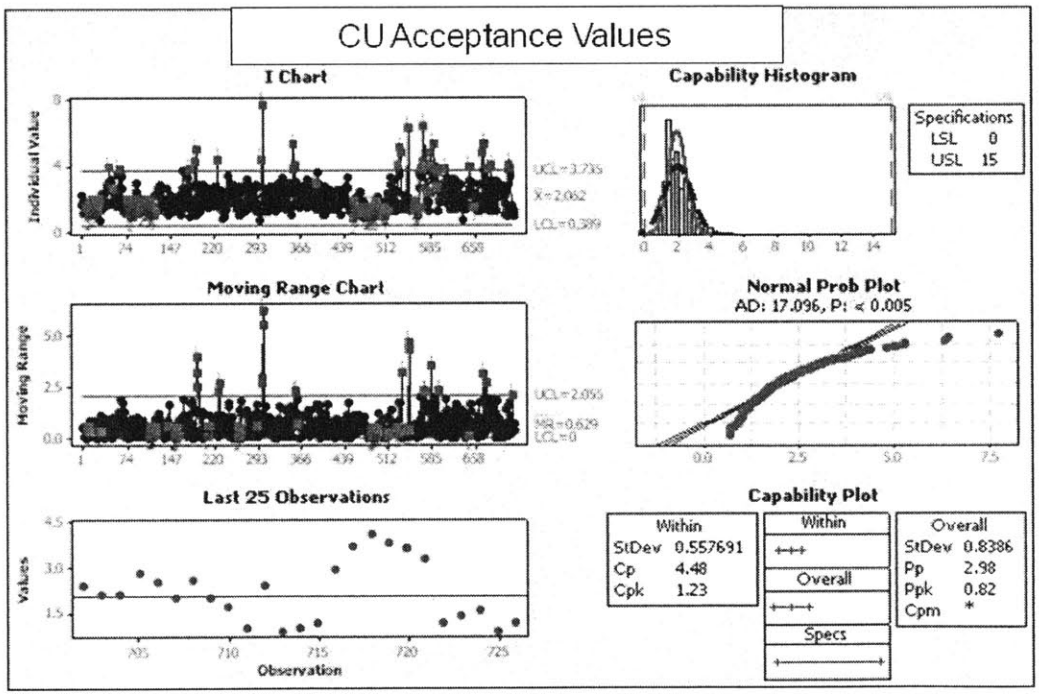

Figure 5. Capability Plot of CU Acceptance Values

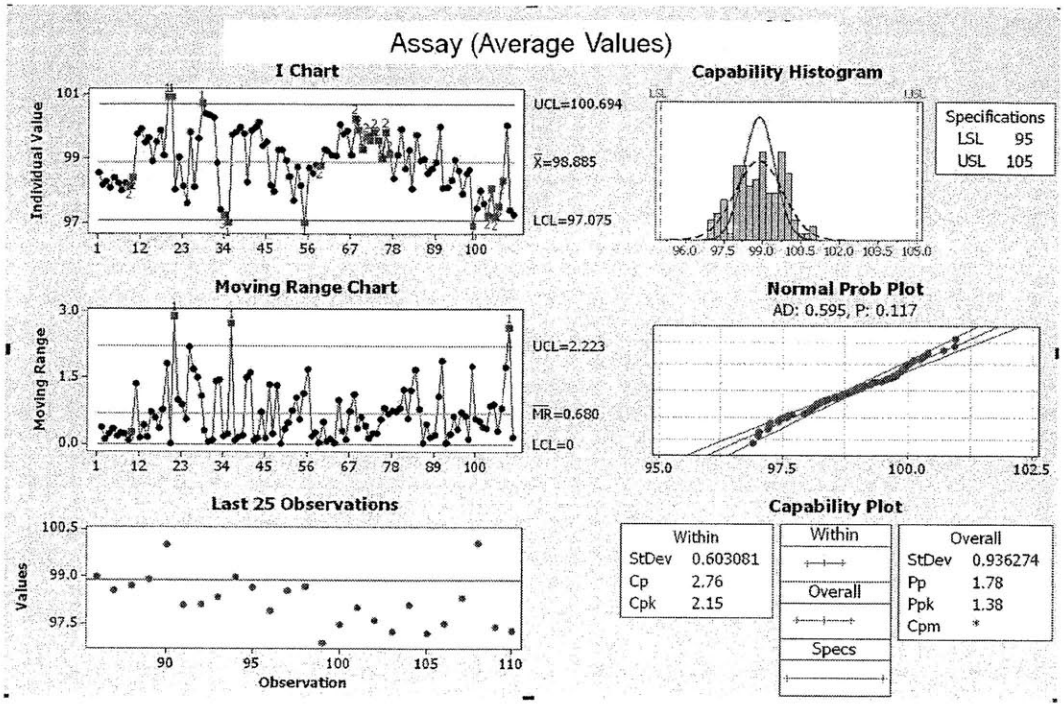

Figure 6. Capability Plot of Assay (Average Values)

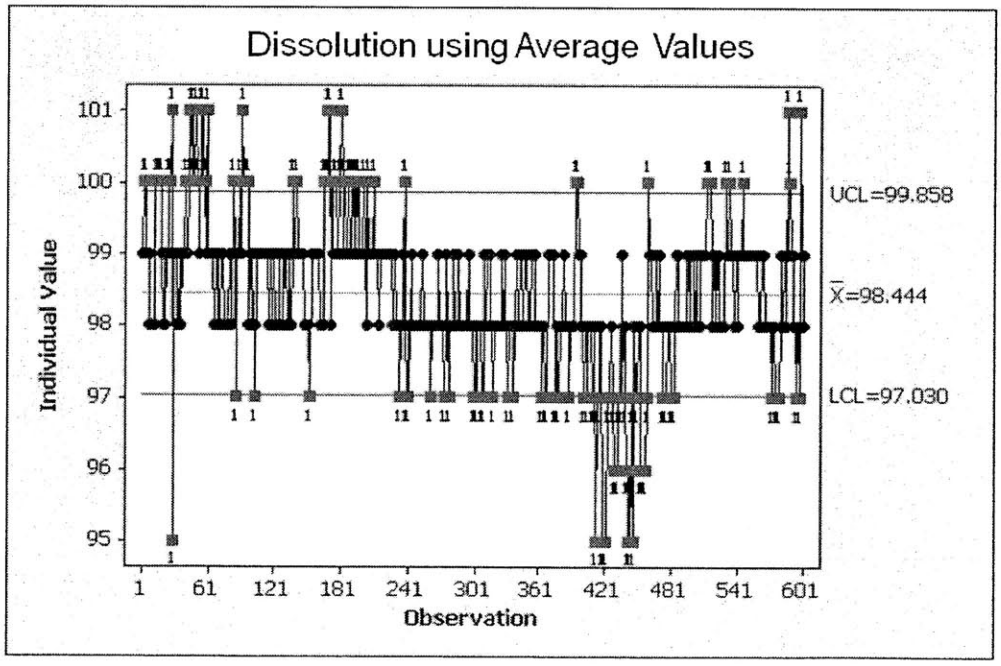

Figure 7. $\bar{X}$ Chart of Dissolution (Average Values)

4.4.1 Process Stability

Viewing the graphs shown above, it is clear that all of them fail the test for process stability. Process stability (as defined in the internal SOP) requires the presence of zero red data points in a chart. In fact a more comprehensive analysis of several data sets from numerous products and attributes reveals almost all data sets when viewed on an annual basis are statistically unstable. Shortening the time frames can improve stability but not substantially.

4.4.2 Normal Distribution

Statistical tools such as process capability and even charts such as the $\bar{X}$ chart (which show the ±3 sigma limits) require the data displayed to be normally distributed. Normally distributed data can be solely described by the arithmetic mean μ and the standard deviation σ. These parameters can be estimated by the sample mean ($\bar{X}$) and sample standard deviation s, respectively. The Central Limit Theorem tells us that the sample size is sufficiently large, then the mean of a random sample from a population has a sampling distribution that is approximately normal, regardless of the shape of the distribution of the population [13]. However, there is no magical number of samples at which point we can declare the data to be sufficiently normal. Furthermore, not all critical parameters use samples means (content uniformity uses a different measure called acceptance value). Process changes can also cause instability in the data and make it harder to obtain a normal distribution. Therefore even though the Central Limit Theorem should in theory guarantee a normally distributed data set, the sample size and the other factors described make it difficult to achieve normality in practice. As described in Chapter 3, without a normal distribution, many of the methods discussed for analyzing process capability and control limits fail. Viewing the probability plots in the graphs above, reveals the lack of normal distribution for many data sets. Minitab uses a p-value of 0.05 to gauge whether the data follows a normal distribution[3] but most data sets fail to show normal distribution regardless of how normality is measured. Adjustment of the time frame can help in obtaining a normally distributed data set but there is no guarantee of this [30].

4.4.3 Independence

Data sets must be tested for auto-correlation to ensure independence. If sample averages are used to represent each batch, we must ensure that successive sample averages are not being

[3] There is no universal minimum p-value which determines normality. Generally lower p-values indicate that the distribution is more likely to have been sampled from a normal distribution.

calculated from the same batch. In cases where batches are re-sampled (additional samples are drawn from the same batch), the data has to be modified to ensure that both the old and new sample data points are excluded. With the exception of a few short data sets that showed high auto-correlation, most data sets showed independence. Short data sets in general are problematic for a number of reasons described later in this chapter.

4.4.4 Time Frame

It is important to note the time frames for the data sets in the graphs. There is currently no active continuous monitoring of quality data, and investigations are only triggered if a drug batch fails to meet the criteria stated on the testing monograph. The red points represent data points that are in violation of one or more of the four statistical tests included in the SOP. However, since no statistical evaluation is performed during the period, these points are rarely investigated unless they happen to fall outside of the internal release limits or the customer specifications. Furthermore, the $\pm 3\sigma$ upper and lower bounds are not actually used in daily practice. Therefore the prevalence of outliers or trends is not surprising. Since there is no monitoring of data based on the guidelines in the SOP, the only method for investigating outliers or trends is to retroactively search the process change record and the database which contains root cause investigations, to look for potential shifts in the data due to material changes or process changes. This retroactive look at data can be misleading because it is not always obvious whether a documented physical change actually caused a change in the data. Unfortunately without active real-time monitoring this is the only available method. Splitting data sets into smaller segments based on physical changes in the process (such as new equipment or a new raw material supplier) can improve both process stability and normality. This bodes well for the online statistical platform proposed in this thesis since it will help in both these areas. However, this brings up some important issues that must be determined prior to implementation of the online system. Are statistical stability and normality absolutely necessary to assess process performance? Are control limits based on ± 3 standard deviations from the mean appropriate with these data sets? Resolving these questions is important for overall success because despite new sophisticated software tools and near real-time automated analysis, the data will remain the same.

4.5 Resolving Non-Normally Distributed Data

Most charts including $\bar{X}$ charts and process capability calculations require data to be normally distributed. The symmetric control limits seen on charts results from the assumption of normality. In the case of highly skewed distributions, a typical $\bar{X}$ chart would provide very misleading results [20][21]. As stated previously an $\bar{X}$ chart is an ongoing test of the hypothesis that the process is operating in a state of statistical control [19]. If a symmetrical $\bar{X}$ chart is used for a data set that has a skewed distribution, the probability of both a Type I error and a Type II error will be different from those which are typically associated with a $\bar{X}$ chart. For example, if a process is within six- sigma, the α-risk (the risk of committing a Type I error and concluding that the process is out of control when in fact it is stable) is 0.0027. However, if the actual distribution of the sample values is not normal, the risk could be much higher. Borror, Montgomery and Runger [22] studied the behavior of the Shewhart control chart for process data that is not normal. They concluded that even if the process shows evidence of moderate departure from normality, the control limits given here may be entirely inappropriate [13].

4.6 Reasons for Non-Normal Data

In this section we will cover some of the reasons why a data set may not be modeled using a normal or Gaussian distribution.

4.6.1 Data is close to zero or a natural limit

A data set may be skewed towards one side of the mean because a natural process limit exists. For example zero is often the natural process limit when describing cycle times and lead times. Since time cannot be negative, a distribution may have a large number of points close to zero and a long tail to the right as shown in Figure 8.

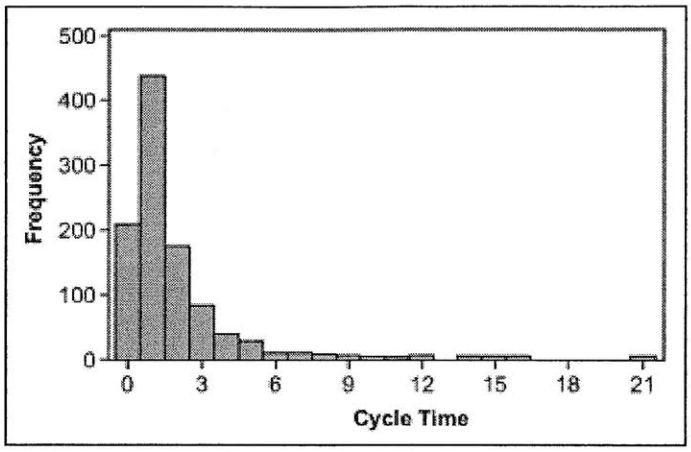

Figure 8. Example of Non-Normality due Natural Limit at Zero [10]

4.6.2 Extreme Values

Some data sets will appear to have a skewed distribution due to too many outliers that are far from the mean. It is possible to fix this problem be reanalyzing the data and removing outliers that have a known root cause. As discussed in the section on normal distribution, small percentage of extreme values should be expected, so only points with known explanations can be removed.

4.6.3 Multimodal Distribution

Process data distribution may appear to be non-normal because it is actually the combination of two or more different processes. These data sources may be normally distributed on their own but when overlapped the combined data may look bimodal or multimodal. For example, two pH testers in a lab may not be calibrated using the same solution. Therefore each tester may produce data that is normally distributed but the mean pH from one tester may be different from the other tester. The result will be the superposition of two normal distributions with different means.

4.6.4 Data Shifts

Other potential reasons for not obtaining a normal distribution revolve around significant shifts in the data. A time series plot that shows large shifts in data or seasonal fluctuations will lead to a bi-modal distribution and or a large variance. If these shifts can be identified, the data should be split into two sets and analyzed separately.

4.6.5 Individual Values vs. Average Values

All the values used for dissolution and assay are average values from batch samples. If the individual samples come from a batch that should be normally distributed, then the averages across batches should in theory also produce a normal distribution [13]. Therefore the use of average values rather than individual values should not cause non-normal distributions as long as the number of samples is sufficiently large. With the new statistical platform proposed in Chapter 5, analysis can be performed on both individual values as well as average values.

4.7 Methods for Handling Non-Normal Distributions

There are numerous methods for handling non-normally distributed data. Identifying a mathematical distribution can help to develop alternate control limits. Many software packages today (including Minitab) have algorithms to help identify whether data belongs to a particular class of distributions. If a data set fails to fit a particular distribution, various mathematical transformations can be applied to obtain a normal distribution. Another method can be to analyze a histogram of the data and use percentiles to set control limits. All these methods are analyzed below. It is important to note here that charts based on samples from a non-normal distribution will have asymmetrical control limits [23].

4.8 Transformations

As stated earlier, non-normally distributed data poses a number of problems for statistical analysis. Over the years a number of mathematical transformations have been developed to overcome this problem. Transforming data means performing the same mathematical operation on each piece of original data. A common example of a transformation is converting miles into kilometers by multiplying the data by a constant (in this case 1.609). This type of transformation is known as a linear transformation because it scales the data but will not change the overall shape of the data distribution. In linear transformations, data is simply multiplied or divided by a specific coefficient or a constant is subtracted or added. Since the actual shape of the distribution remains unchanged, linear transformations are not very useful for helping data look more normal. A more useful transformation to achieve normality is the Box-Cox transformation.

4.8.1 Box-Cox Transform

The Box-Cox transformation is a procedure that helps identify the optimal power transformation for achieving normal distribution. A power transformation involves raising all data to the power of lambda. For example, if all data points are squared then lambda will equal 2. In 1964, statisticians George Box and David Cox developed a procedure to identify an appropriate exponent (lambda) to use to transform data into a "normal shape" [24]. The Box-Cox algorithm searches for the optimal lambda (between -5 and +5) that will transform the data into a normal distribution. However, there is no guarantee that the optimal lambda will produce a normal distribution. Once the optimal lambda is selected (one that is most likely to yield normal distribution), the transform must be applied to the data set, and probability plot must be performed.

The Box-Cox transformation is defined by the following piece-wise function:

$$x(\lambda) = \left(\frac{X^{\lambda} - 1}{\lambda}\right) \quad \lambda \neq 0$$

$$x(\lambda) = \ln(x) \quad \lambda = 0$$

At λ=0, the function would produce a singularity since any value raised to the power of zero yields a value of 1. By defining a separate equation at λ=0, $x(\lambda)$ becomes a continuous function.

There are several methods for obtaining an optimal lambda such as finding the lambda that maximizes the log-likelihood function. However, the most straightforward approach is to choose a lambda for the Box-Cox linearity plot so that it maximizes the correlation between the transformed x-values and the y-values when making a normal probability plot of the transformed data. Though this brute force method is more CPU intensive, today`s computers can easily handle this method even with large data sets. The output from this method will also yield the p value for the optimal lambda and indicate whether Box-Cox transformations are appropriate for a given data set.

4.8.2 I-Charts and 3 Sigma Limits using Transforms

The Capability Plots in the Quality Toolkit for Minitab allows the user to automatically apply Box-Cox transformations to data sets. The software automatically determines the optimal lambda (between -5 and 5), applies the transform, and graphs the data. Since the normal probability plot is included, the user can clearly see whether the transform was successful in making the data more normal. The major issue with using transforms (including Box-Cox) is that most users have trouble making sense of the data. See the example below, which shows data collected for a particular QA attribute.

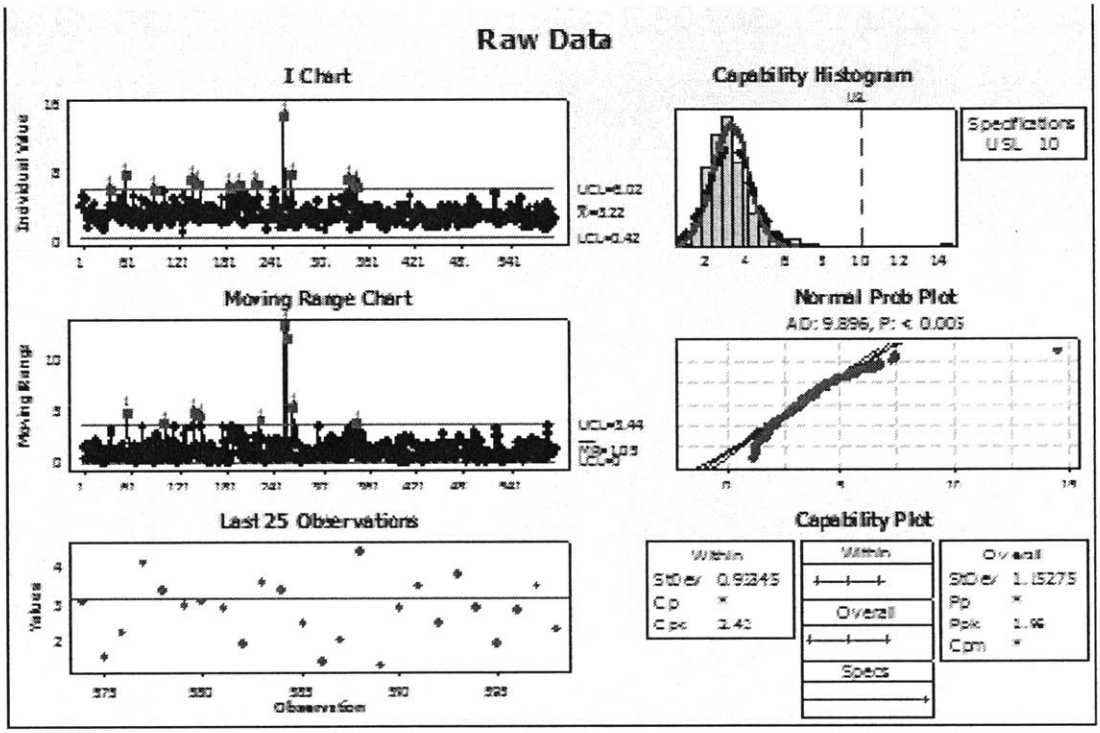

Figure 9. Capability Plot for QA Attribute With Non-Normal Distribution

Though the histogram gives the impression that the data has a bell shape, the normal probability plot indicates that the data is not normal. Visually the data appears to be non-linear on the plot, and the p-value for the plot is less than 0.005. Without a normal distribution the capability statistics (i.e., standard deviation, process capability) do not carry any meaning. If we use a Box-Cox transformation, the optimal lambda is determined to be 0. This is a natural log transformation.

$$x(\lambda) = \ln(x) \quad \lambda = 0$$

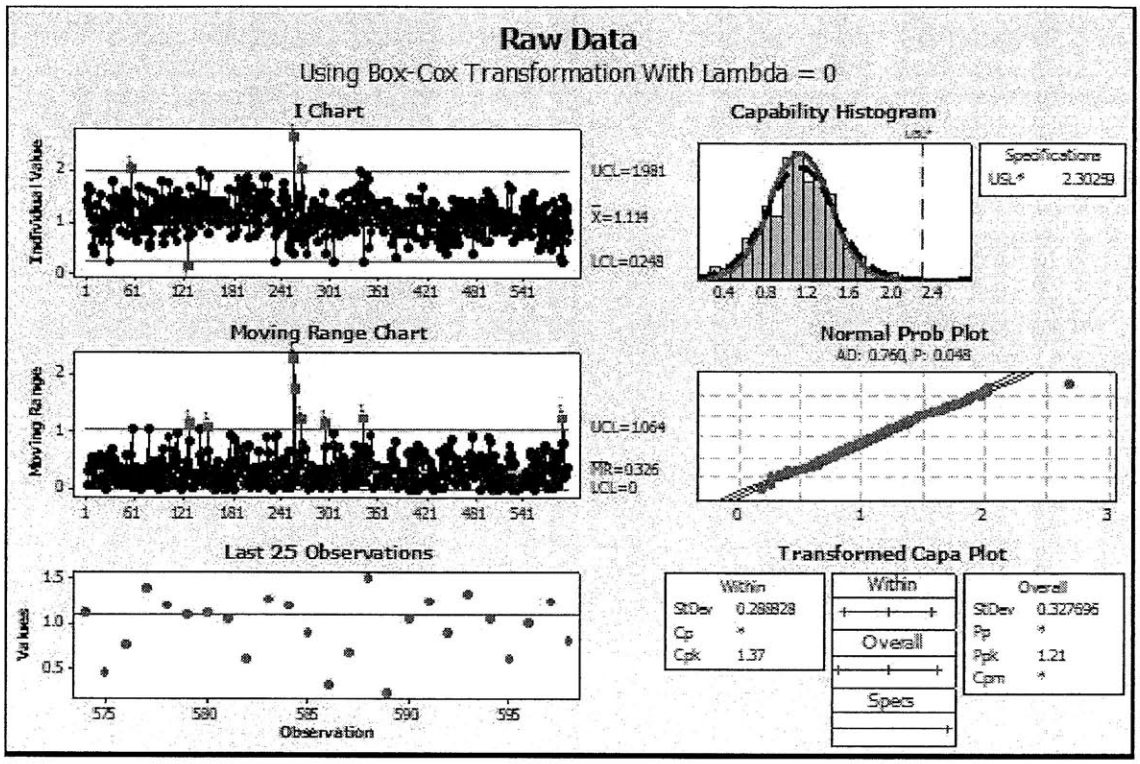

Figure 10. Capability Plot of Data with Box-Cox Transformation

The graphs for the transformed data are shown in Figure 10. The normal probability plot shows a fairly linear plot and a p value of 0.048[4]. The capability histogram uses the transformed data. The USL or upper specification limit has also been transformed as ln(10) = 2.309. The I-Chart and moving range chart are also in the transformed domain. This can be jarring because the numbers have no connection the actual physical process. Conceptualizing the natural log of a measurement can be challenging. To make sense of these charts, the UCL and LCL limits need to be transformed back into the original values. Since $\lambda = 0$, the inverse natural log will transform values back to the original domain.

$$UCL_{Original} = e^{1.981} = 7.2$$

These values can be used to establish the six sigma limits for the original data. A run chart with these values and the customer specifications is shown below. The UCL value now makes sense

[4] Typically a p-value of 0.05 is used as a threshold to determine whether data is normal. In this case 0.048 is very close to this threshold value and we have considered the data to be sufficiently normal.

since it is in the same domain as the original data. These statistical control limits could serve as the internal control limits for this process. Any points outside these control limits would cause an excursion and require a root cause investigation.

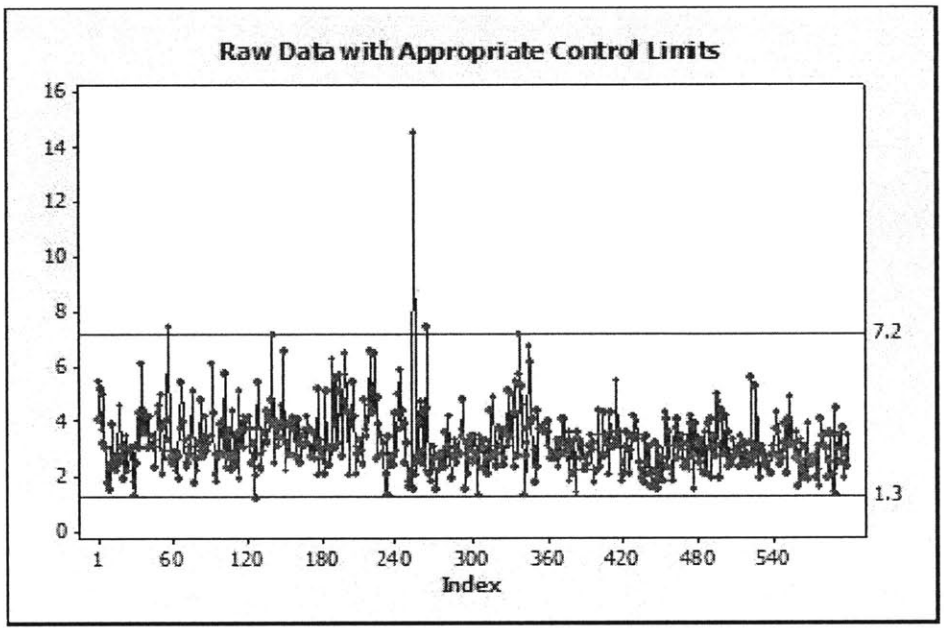

Figure 11. Control Chart with Assymetrical Control Limits[5]

4.8.3 Johnson Transform

Another powerful transform that can help produce normal distributions from non-normal data sets is the Johnson transformation [33]. The transform is named for Norman L. Johnson who in 1949 provided a system of transformations to transform non-normal data to a normal form. The Johnson system is comprised of three curves, Bounded, Log-Normal, and Unbounded. The method by which the algorithm estimates the optimal curve for a given distribution is beyond the scope of this thesis [25] [26]. Modern software packages have no trouble performing Johnson curve fitting, and Johnson transformations are included in the Minitab software package. Using this class of transforms we were able to transform eighty percent of analyzed data sets into approximations of the normal distribution. It is therefore more powerful than the Box-Cox transformation but can used in an identical manner.

[5] Ignore the lower limit of 1.3 in this chart. It is a result of another analysis and does not reflect the transformation.

4.8.4 Non-Normal Process Capability Indices (Percentile Method)

As described earlier, process capability indices are generally computed to evaluate the quality of a process, that is, to estimate the relative range of the items manufactured by the process (process width) with regard to the engineering specifications. For the standard, normal-distribution-based, process capability indices, the process width is typically defined as *six-sigma*, that is, as ± three times the estimated process standard deviation. For the standard normal curve, these limits (z_l = -3 and zu = +3) translate into the 0.135 percentile and 99.865 percentile, respectively. In the non-normal case, the three sigma limits as well as the mean (z_M = 0.0) can be replaced by the corresponding standard values, given the same percentiles, under the non-normal curve [27] .

Shown below are the formulas for the non-normal process capability indices:

$$C_p = (USL-LSL)/(U_p-L_p)$$

$$C_{pL} = (M-LSL)/(M-L_p)$$

$$C_{pU} = (USL-M)/(U_p-M)$$

$$C_{pk} = Min(C_{pU}, C_{pL})$$

In these equations, M represents the 50'th percentile value for the respective fitted distribution, and U_p and L_p are the 99.865 and .135 percentile values, respectively, if the computations are based on a process width of ±3 times sigma. Note that the values for U_p and L_p may be different, if the process width is defined by different sigma limits (e.g., ±2 times sigma). For more on the use of percentiles see [28].

4.9 Statistical Process Control Plan in an Ideal Setting

Lack of process stability and normal distribution pose serious problems for developing charts and calculating process performance. The ideal plan summarized in Figure 12, would be to set the thresholds based on three sigma limits, monitor the process and investigate any points that fall outside this limit. Once sources are determined, the process can be modified to remove these sources of variability. The data set can re-evaluated annually to determine the new three sigma limits and over time the process will improve. Once process stability and a normal distribution are achieved, process performance can be calculated and used as a metric. This plan meets all the

requirements for statistical process control and will eventually lead to process improvement. Unfortunately there are some major drawbacks with this plan.

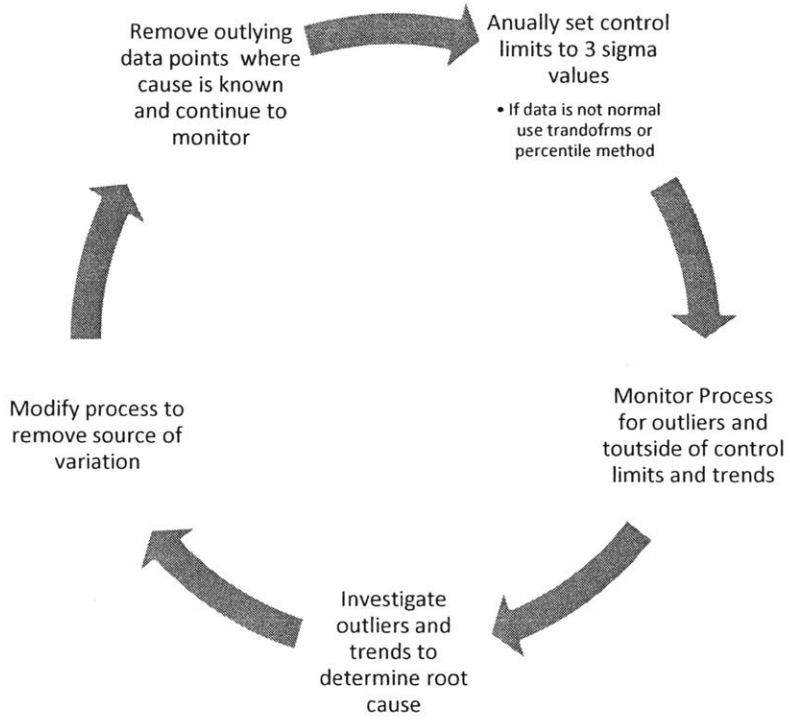

Figure 12. Process Improvement Cycle

4.9.1 Productivity

Setting the control limits to the three sigma limits would generate a large number of deviations requiring root cause investigations. Even if the new plan is piloted for one product campaign and limited to a few processes, the exercise could be quite taxing on the organization. A new statistical software platform, with more advanced tools, dashboards and collaborative tools will help but the productivity gains achieved would be erased by all the new root cause investigations.

4.9.2 Warning with Transforms

One of the problems with the use of transforms when applied to our data is that they tend to be inconsistent when the time frame is enlarged. If a transform is robust and correctly describes the underlying distribution, then the transform should not change substantially even when the time frame is changed. However, for many of the data sets, we found that when the

time frames for a particular attribute changed, the transform was completely different indicating it is not very robust (that is, it is sensitive to the subtle stability changes in the underlying process).

4.10 Recommendations

Based on the data collected and literature review, it is felt that normal distribution and process stability are important for calculating a valid process capability index. However due to the difficulty in achieving these requirements, there needs to be some flexibility surrounding these assumptions. This is a compromise we must make to account for the fact that almost no process that was reviewed met both these requirements. Novartis IQP has in fact agreed to this change and has issued a new guide chart for instances where one or both assumptions remain unmet. The guide is a general document for all processes. In the section below, we provide further recommendations for the three critical parameters reviewed as part of this project.

4.10.1 Content Uniformity

As discussed in Section 4.3, content uniformity is a measure of how consistent a product is in terms of the percentage of active ingredient. It already has statistical analysis incorporated and this is why it should not be surprising that it does not follow a normal distribution. In fact, depending on the target range for CU, the acceptance value for CU is likely to be proportional to the standard deviation of the samples. Recall,

$$AV = \left|M - \overline{X}\right| + Ks$$

When we plot the difference between acceptance values (AV) and Ks values as shown in Figure 13, we find the differences to be small. Since Ks is based on the standard deviation, we can conclude that AV is proportional to standard deviation.

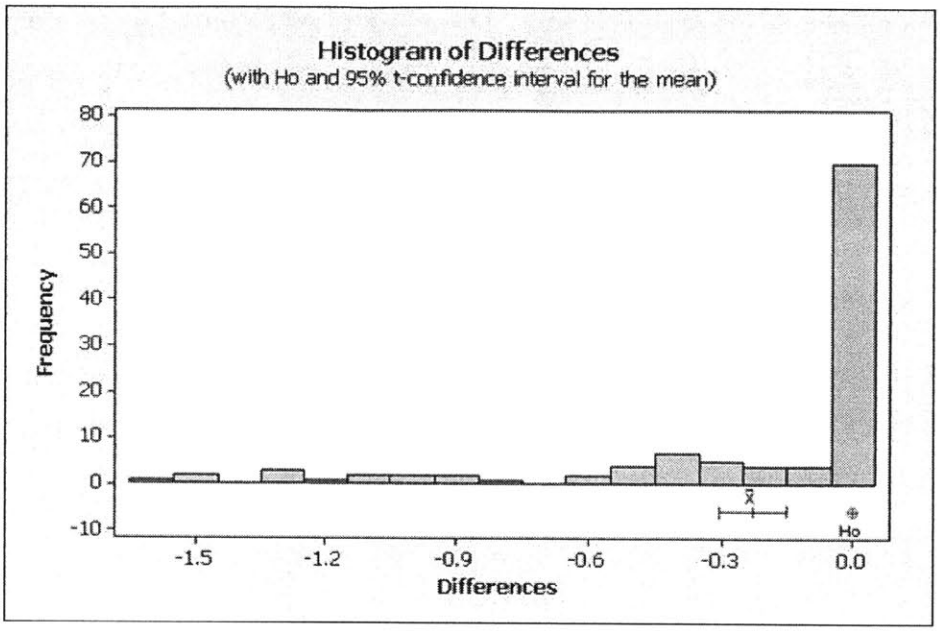

Figure 13. Graph comparing Acceptance Value with Ks

We should also consider the fact that the control limit for acceptance value is 15, and most of the values we have are considerably lower than this. With this in mind, it is proposed that in cases where there is no normal distribution or process stability, a simple acceptance chart should more than suffice for this attribute. An acceptance chart is simply a time series chart that has horizontal lines displaying the customer specification limits. It is not a control chart since it simply tells us whether a data point is acceptable (in-specification). Now it is true that an acceptance chart lacks more sophisticated metrics such as process capability ratios. If we were to use transforms we could in fact obtain a normal distribution in many cases and calculate process capability. However, the percentage of cases in which a transformation is successful is not high enough to justify the endeavor.

4.10.2 Assay

Assay is a key critical parameter and must be monitored closely. In cases where there is no normal distribution, a transformation is recommended for establishing control three sigma control limits. The use of Johnson transformations, for example, should be able to produce normal distributions and therefore allow us to calculate process capability. As with other processes, the effect of process changes must be monitored to ensure that data sets for different processes are not combined. Therefore the time series chart may be split into many sub-segments. In these instances, if the sample data set is too small, we are unlikely to obtain a

normal distribution required for calculating process capability. Even the proposed IQP guidelines will not suffice because the distribution may be very far from normal. Therefore in these instances a run chart will have to suffice for monitoring until enough data is collected for us to use a transform to obtain normal distribution. Once these limits are established, they do not have to be recalculated until the process shifts or the control limits are changed.

4.10.3 Dissolution

As can be seen from the dissolution data shown in Figure 14, the discrete levels pose an issue for achieving a normal distribution. A plot of mean sample values leads to quantization and makes it virtually impossible to achieve a normal distribution. A plot of all individual data points solves this first issue.

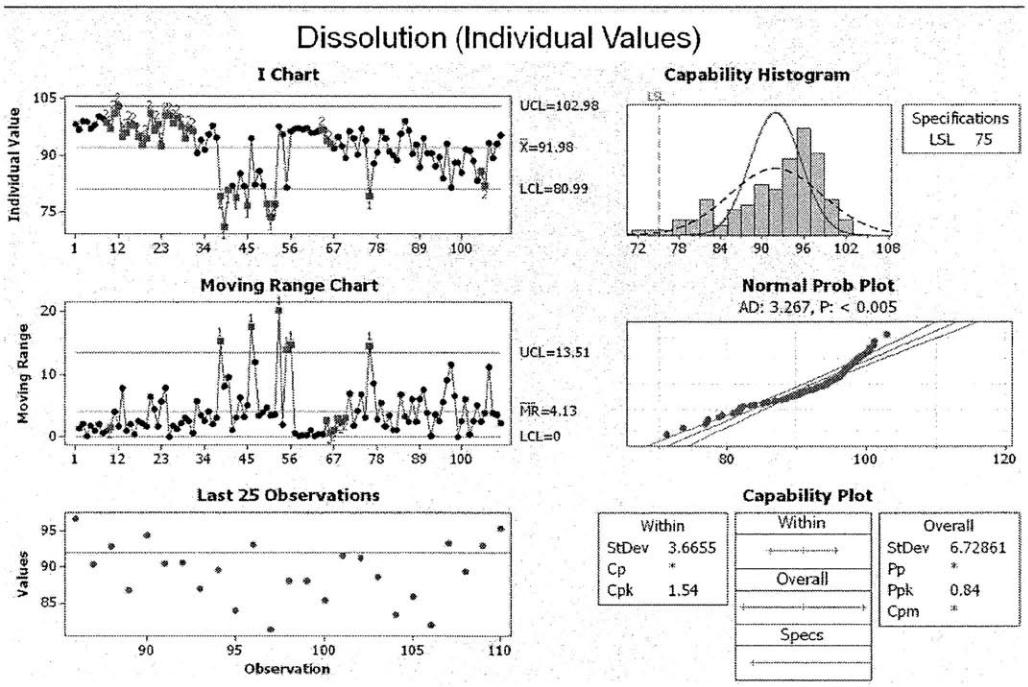

Figure 14. Plot of Dissolution Values (Individual Values)

However a chart with individual values (like the one shown in Figure 13) is still unlikely to reveal a normal distribution. In this case, we would recommend the use of the *percentile method* for setting the three sigma control limits. To set the control limits using percentiles, we graph the

data set in the form of a histogram and then calculate the area under the curve and set control limits equal to the area under the tails of the corresponding three sigma limits. The one cautionary note here is that since individual values are used (and not sample means) we risk violating the independence rule. Violating the independence rule may lead to too many false alarms [13] and if this does occur, the procedure may have to be revised. It should also be noted that there are other tests that can be performed using individual values. For example, we can calculate data range values, as well as standard deviation values for each batch (based on the individual sample values) and use this to measure intra-batch variability. However, for the purpose of ensuring that dissolution remains above the minimum specification limit, a histogram using six sigma limits based on the corresponding percentage values will be most useful.

5 Decision Support System

In this chapter we will explore the databases and software packages currently used to generate the APR/PQR report and the corresponding statistical analysis. After a brief summary of the current architecture, the chapter focuses on the tools and capabilities needed to automate the report generation process and increase productivity. Alternate software packages for statistical analysis are evaluated and the software tool *Signal From Noise* is recommended as an alternative to *Minitab* for the automated system.

5.1 IT Architecture

The current IT architecture for Quality Assurance /Quality Control consists of a myriad of databases and software packages. Many of these databases are not directly linked, and this makes data consolidation a time consuming task. In fact in preparation for the APR/PQR, an employee must first consolidate all relevant data from the different databases into a separate spreadsheet and then e-mail this spreadsheet to the various parties who will assist in assembling the document. A few of the relevant databases are listed in Figure 15.

Database	Relevant Information
Chromeleon	HPLC QC Data
SAP QM	Central Database for QC Data
TEDI	Stability Data and recalls

Figure 15. Database description.

Recall the goal of the APR/PQR automation project is to support the creation of the APR/PQR reports as well as tools for online quality monitoring and trend analysis. This process involves a number of different steps which are discussed in the subsequent sections.

5.1.1 ETL (Extract, Transform, Load)

Extract Transform, Load is the initial operation required to extract data from a database, perform initial transformations on it if needed, and then store it in a data warehouse. Initially a data extraction tool must extract the specific data required for a particular APR and store it in the Data Warehouse (DWH). The extraction tool requires specific data pumps to be programmed with different databases in order to perform optimally and therefore there is a need to reduce the number of databases for extraction. As a first step, it is proposed that SAP QM be the central database for all QC data that is gathered at the QC labs in Stein. This requires building a data pump from Chromeleon to SAP QM which will eliminate the need for users to manually enter data which is both cumbersome and time consuming.

5.1.2 Data Warehouse

A data warehouse is defined as "a place where data is stored for archival, analysis and security purposes" [29]. The idea behind data warehousing is to create a database where all data from different databases can be consolidated and transformed for further analysis. One might ask, should SAP QM simply not be made the data warehouse for this system? There are a number of reasons this solution will not suffice. The APR/PQR report consists of data beyond that which is stored in SAP for example, drug stability, deviations, batches produced, and drug complaints data. These data types are kept in separate databases and integrating them into SAP could be cumbersome. Another important reason for creating a separate data warehouse for QC data specifically concerns the use of data transformations. In order to maintain data integrity, SAP QM should store raw data collected from the HPLC tests performed on the lab bench. Later these values may have to be modified (for example rounded to the nearest integer) before analysis can be performed on the data. For data integrity reasons it is important that data not be modified in any manner in SAP QM. Therefore any modification of data should occur after SAP QM, and a separate warehouse provides an ideal location for this operation. This is why the ETL operation will be performed between SAP QM and the data warehouse.

Another important parameter that must be determined is the frequency of replication between SAP QM and the data warehouse. This is highly application specific, but given that lab tests are generally performed once a day and the lab technicians are only available on day shift,

replication does not need to occur more than once a day. Therefore the data ware house is an *offline* data warehouse. In the future if in-line real-time process data is incorporated then the replication frequency may have to be adjusted.

5.1.3 User Content Management System

The user content management system is responsible for selecting data from the data warehouse and grouping it into data objects. After creating objects, UCM will import these data objects into a document management system that will generate the APR/PQR template and fill in the appropriate fields. The document management system will use Microsoft Word for building the templates and content formatting. After all the edits of the Word document are complete, the report will be converted into an Adobe® PDF file for approval and finally into an electronic record for submission to health authorities. The process for completing the APR/PQR is described below.

5.1.4 Workflow Management System

The Workflow Management System (WFMS) handles the lifecycle of the APR/PQR report. Typically, the reports go through multiple levels of edits and approvals and it is important that this "workflow" is monitored and secure. The WFMS will first trigger the creation of a report based on annual reporting cycle for drug registration. The trigger will send an email (or some other message) to the QA staff member alerting them that a product's APR/PQR report is now ready to be created. The UCM will automatically generate the Word document and populate the document with appropriate data objects. Several individuals may collaborate on the report by working on different chapters. The system will handle the collaborations as well as multiple versions of the document. Versioning is useful when different collaborators need to check out a document, make changes or retrieve a previous version of the document. Digital signatures will be used to handle the approval process and if an edit needs to be made at this stage, the document will be released back to the collaborators who will edit the document, and resubmit. All changes will be tracked by the WFMS so that a clear audit trail for the document exists. Once the approvals are granted at the plant level, the document will be released to the Country Pharma Organization (CPO) for release of the respective product to the market. Finally the system will also provide archiving and record management so that reports are easy to search.

5.2 Business Process Intelligence

The various systems described in Section 4.1 together comprise the overall Business Process Management (BPM) for the APR/PQR generation. What the BPM is missing, is a method for creating and displaying meaningful analysis from the extracted data. In the IT industry, this process is termed Business Intelligence (BI). In the APR/PQR scenario, the BI tool should graph the data, identify trends and point users to critical areas that deserve the most attention. Since this information is useful to both APR/PQR generation, as well as everyday quality monitoring, the BI tool should also include a dashboard for general quality tracking. A dashboard is essentially a web page that uses a number of dials and colors to give users a quick glimpse of a process and alert users to how well various processes are performing. The data used to generate the graphs and analysis for the APR/PQR can also be used here to run the algorithms that monitor quality performance. To satisfy this, the software selected for BI should be flexible enough to generate graphs and export them to the document management system as well as display performance metrics on a dashboard. There is some controversy as to whether one tool can handle both these functions. An alternate approach would be to select one software system for the APR/PQR, and have a separate one handle the statistical analysis for the dashboard. After conducting industry research on the various software options available and speaking with a number of vendors, it appears there are several software packages that can perform both functions. These software packages are compared with the software currently used for business intelligence in Section 5.4.

5.2.1 Vision for QA Dashboard

Before software packages are evaluated for the BI platform, the vision and requirements for the dashboard must be clarified. At a high level, the dashboard must convey information about how well the manufacturing process is performing from a QA perspective. Users should be able to view the dashboard and quickly identify critical quality issues that deserve the most attention. At the same time, the dashboard should also provide information useful for performing root cause analysis. Finally, the dashboard should be an extension of the APR/PQR Business Process Management system. The dashboard should allow users to continuously monitor the same parameters and graphs that will eventually make up the APR/PQR report. This ensures that

issues are detected early and are corrected, instead of waiting till the annual review period is over. In an ideal scenario, the dashboard should help build a more stable and capable manufacturing process. In order to do this, a system must be designed that will automatically make the charts, alert the user to data that violates the predefined rules, and update process capability in real-time. A potential layout for the dashboard is shown in Figure 16.

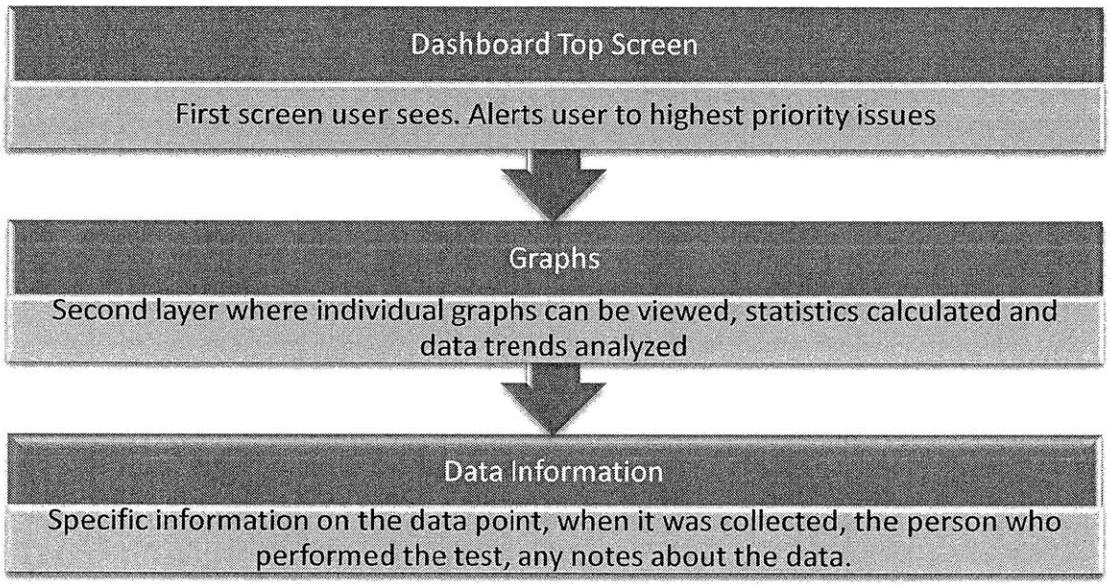

Figure 16. Dashboard Hierarchy

The opening layer of the dashboard provides a high level snapshot of the manufacturing process and identifies deviations that violate user-defined rules. The screen should provide an easy to grasp visual representation of process capability so that users can immediately identify which processes need improvement. Color coded dials are one method of representing this information. Each critical parameter will have a performance dial indicating performance. Process capability values for various parameters can have pre-defined ranges for red, amber and green. The diagram in Figure 17 illustrates this.

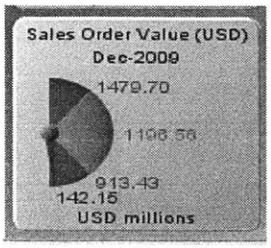

Figure 17. Dashboard Meter Dial

The dashboard should also provide some detailed information about deviations so that users know what issues need to be investigated. Issues requiring potential follow-up are identified based on the statistical rules defined in Chapter 2. An example of this type of information is provided in Figure 18.

Date	Product	Attribute	Rule Violation
2/3	Drug A	CU	3rd consecutive point Out of control limit
2/2	Drug B	Assay	9th consecutive point above mean

Figure 18. Example of Dashboard Alerts

The dashboard should also have a method for inputting comments into the system based on root cause investigations. For example, say a process deviation is investigated and it turns out that the root cause is a switch in raw material suppliers. The user should be able to input this comment into the data point associated with the deviation so that anyone else opening the graph will see that this deviation has already been addressed. Furthermore, these comments should be stored as objects and imported into the APR/PQR reports along with the graphs. By offering this feature, the user can simply keep the dashboard open at all times, rather than juggling multiple databases simultaneously.

5.3 Evaluation of Minitab

The software package currently approved by IQP for statistical analysis at Novartis Technical Operations is Minitab Version 15. In fact the procedure for performing process capability analysis of Quality Assurance Data is titled Process Capability Analysis using Minitab® Statistical Software. While the software is able to perform the tasks required for process analysis (distribution chart, calculation of mean, standard deviation, etc.) it does pose some shortcomings when integrated into an automated IT solution. Minitab is an adequate tool for detailed statistical analysis in one-off applications, but it lacks many of the capabilities required for an automated IT solution. In an automated solution, data should automatically be pulled from SAP QM and analyzed with built in algorithms. The results should be either displayed on a dashboard or automatically exported to the appropriate APR/PQR report. Minitab provides connections to other databases using data pumps such as ODBC[6], but once data is downloaded to the program it cannot be protected. Creating the six pack analysis (required as part of the SOP for process capability) can be automated using macros, but the displays cannot be automatically exported to a third party platform. The use of Minitab to build a dashboard is technically feasible and the support staff at the software vendor mentioned a company that is using the software in this fashion. However, the example is very complex to implement and would not serve as an acceptable solution[7]. Therefore other software has to be investigated in order to determine its feasibility for the project.

5.4 Criteria for Statistical Software Selection

Before evaluating alternate software packages, it is important to have a set of criteria. Based on the requirements of the Global APR/PQR Automation project and the shortcomings identified with the current Minitab solution, new criteria are required for evaluating competing software solutions. The quantitative ratings for these criteria are subjective and based on the features of each software package.

---

[6] Open Database Connectivity (ODBC) provides a standard software interface for accessing database management systems (DBMS)

[7] Project Meeting, Novartis Basel, Switzerland September 10th, 2010

5.4.1 Data Security

Data security refers to the mechanisms for protecting data and ensuring data integrity. The number of instances that data can be manipulated between the instant the test is performed to the time of analysis must be minimized. If any changes to the data must be made, these should be done in the actual database where the data is stored (such as SAP QM) and access to this database must be restricted. Examples of this include password protection and sum checks. Users should not have the ability to modify or delete values when importing data from the data warehousing database into the analysis tool.

5.4.2 Flexibility

Flexibility refers to the number of different tools available within the software package for statistical analysis. Most packages come with standard features such as run charts, and histograms. However, more advanced tools may be necessary such as distribution identification, and transformations. Another highly desired capability is the ability to split a set of data into stages and display all stages on the same graph. An example of this is shown in Figure 19.

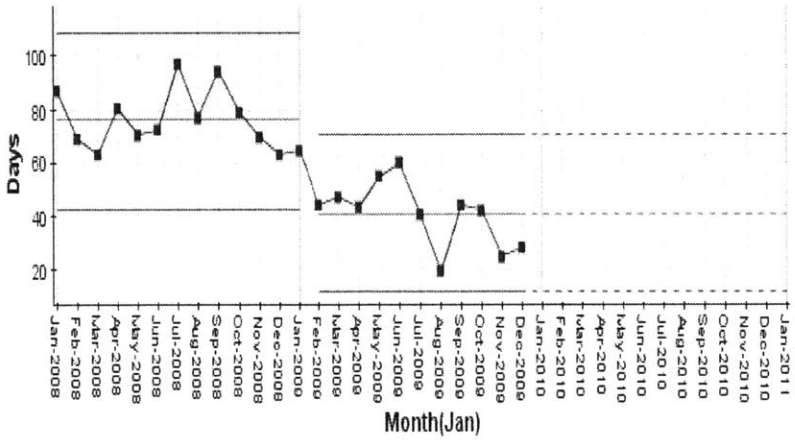

Figure 19. Example of Run Chart with Split Data Sets

Data may be separated into stages and evaluated for performance separately for a variety of reasons. One reason could be if a major process change is introduced and this change is known to impact the measured data (for example by shifting the mean). In this case, the data after the change should be analyzed separately so that the impact of the change can be accurately

measured and not diluted by the previous data. This feature is vital for analyzing process performance and measuring the impact of changes.

5.4.3 Automation/User-Friendliness

This automation and user-friendliness category refers to the ease with which tools can be accessed and the level to which routine analysis can be automated. Minitab, for example, allows customized macros to be written which helps automate the creation of charts. However, the fact that the program must be separately opened reduces the level of automation when considering the statistical software as part of the overall IT solution. Software that can be integrated into a business intelligence platform seamlessly without the user having to open the software will have a higher rating.

5.4.4 Validation

All software used for statistical analysis must first be approved by Novartis IQP. This group will validate the software by performing a number of tests on sample data to ensure that the results are consistent with the results from approved software platforms. In theory all professional software tools will eventually be validated. However, some software packages are already used in other departments at Novartis and these packages will be validated faster.

5.5 Software Review Results

The software packages selected for review are based on industry research, software already in use by other Novartis subsidiaries and conversations with IQP members who have experience with statistical tools. Several vendors also provided demos for trial use. The results of the evaluation are shown in Figure 20.

	Data Security	Validation	Automation/ User Friendly	Flexibility
Minitab	2	3	2	3
IMIS	3	1	3	2
Signal from Noise	3	2	3	3
SAP Business Object	3	2	1	2
StatGraphics	2	2	3	3
OptimProcess	3	2	3	2

Figure 20. Statistical Software Evaluation Results

The numerical ratings are subjective and based purely on how well the software meets the project criteria. The software vendor that appears to be the best fit for the project is Signal From Noise (SFN) whose vendor is Lightfoot Solutions Ltd. Interestingly many of these packages are in trial use with different Novartis divisions. Optim Process is in the testing phase with Novartis Biologics in Huninge, France while SFN is being tested for use in supply chain and operations in Stein, Switzerland. IMIS is a software package currently being developed in-house with support from multiple Novartis sites. Ultimately, SFN provided by Lightfoot Solutions, is the best choice not only because of the evaluation ratings, but also because many of the data pumps used for communicating with SAP are already in use for supply chain applications. This thereby reduces the amount of time for additional programming and ensures that the package will function with QC data sets.

5.6 Review of Signal From Noise®

Signal From Noise provides the ideal BI platform for both the APR/PQR and the dashboard. The software contains layers of analysis very similar to the layers described in Section 4.2.1. The top layer is the actual dashboard made up of multiple dials for tracking the performance of various parameters. Furthermore, the dashboard can be customized for different users so that a user only sees the performance dials relevant to them. For example, one dashboard

can be built for tracking products in the solids department while a separate one can be built for vials and syringes departments. A comprehensive dashboard featuring all dials can be built for the QA Head.

SFN also provides a variety of SPC charts for graphing various metrics and comparing the performance of a metric over different periods of time (second layer of the dashboard model). It allows the user to input a split into a graph if a significant event has occurred, requiring the data set to be split. The software also allows the user to benchmark a process parameter over different production units or even different plants. This can be especially useful for Novartis since many plants produce the same drug and experts may be interested in looking at trends across all plants. The charts allow the user to click on a data point and enter a comment associated with that point. These comments are stored on a separate table in the database and can be linked with the objects for the APR/PQR. SFN provides a key feature concerning the area of data suppression. As discussed in Chapter 3, data outliers with known root cause determinations should technically be excluded from process capability calculations. This is because these points are not representative of the typical process and can skew both the process stability and capability. However, given the importance of quality data in terms of regulatory compliance, management is very cautious about suppressing outlying data points even if these points are known to be erroneous. In a regulatory audit, this could be construed as hiding data from regulators. SFN solves this dilemma by allowing the user to suppress a data point for calculation purposes, but displaying the data point in charts. The data point shows up with an asterisk on it, indicating it is suppressed. A comment will be associated with this asterisk indicating the reason for the suppression and a reference to the root cause investigation.

Drilling down an additional layer, any data point or data set can be analyzed at the "transaction" level. This corresponds to the third layer of the dashboard model where the user can find information about individual data sets such as the date/time stamp, lab equipment number and many other additional attributes. This information is useful for launching root cause investigations for deviations. The software allows the user to select a particular data range and download all the data for this range to an off-line location. This is a powerful feature because the current method for performing this can take several hours and requires access to multiple

databases. These data sets could be analyzed in detail using more powerful statistical software such as Minitab.

6 Organizational Implications

The details of the statistical IT platform and the analytical techniques presented in this thesis alone will not address all the challenges facing Novartis Quality Assurance. To achieve success, the suggestions must be incorporated into both the overall APR/PQR automation project and adopted by the various departments within in the company. Accomplishing this requires us to address some of the strategic, cultural and political challenges within the company.

6.1 Strategic Lens

Novartis is a large pharmaceutical firm with several different divisions. The company has deliberately not forced standardization across all the divisions and has chosen to operate each division as if it were a separate company. Corporate management clearly sees some advantages to this approach including reduced bureaucracy and more flexibility. However, there are some disadvantages to this approach such as lack of knowledge sharing across divisions and duplication. Within the Pharmaceutical Operations division there is greater central control through global standard operating procedures and sharing of information on regulatory inspections. In the context of this project, Corporate Quality Assurance has allocated funding to automate the Annual Product Review and make it easier for departments to consolidate and share data. As described in this thesis there are many mechanisms available for achieving this automation and it is important that both the end users (in Stein) and corporate share the same vision for the tool in terms of functionality and application. If not, we run the risk of the tool not being utilized because the end users fail to see the value in it. This can be avoided by organizing regular meetings between Novartis IQP members and local QC personnel. Procedures for the new tool can also be reviewed and refined by applying the existing guidelines and SOPs to actual data that will be used in an APR/PQR. These meetings will also help tackle scenarios in which the guidelines fail (such as lack of normal distribution) and help ensure that the delivery of the new software tool coincides with procedures that are best suited for the data QC reviews on a continuous basis.

6.2 Cultural Lens

All departments and stakeholders involved with this project share the common goal of meeting the ICH Q10 guidelines and ensuring that the process of producing drugs is

continuously improved. While achieving this vision, the automation project also serves as a symbol to different stakeholders. For the local QA lab technicians it is a tool that will help automate data consolidation and allow for more comprehensive data analysis. The technicians rightfully hope that this tool will make it easier for them to perform their analysis. To the Global IQP champions, the project represents an opportunity to finally bring statistical process control concepts and data monitoring to the local site QA departments. There are however some concerns about the level of automation being incorporated into the final tool. There is some merit to this concern in that someone still needs to perform the analysis of data sets and some manual work will still be required. Some have pointed out that regardless of the level of automation promised through this project, human intuition and analytical rigor is still required to search for trends and determine their sources. They believe that charts printed on paper and signatures still hold value in the digital age. While we may not advocate printing all daily individual graphs, the task of looking at the data and making a digital comment should be incorporated into the daily responsibilities. Otherwise the daily monitoring tool will sit idle and unused.

The automation tool also brings with it an opportunity for local QA/QC, IQP and operations to work even more closely on root cause investigations. Many of the lab technicians have a pharmacy or chemistry background as opposed to IQP champions who tend to be engineers. Therefore lab technicians are trained in running experiments, and following detailed procedures, while IQP is perhaps more data driven and accustomed to using analytics to discover problems and fix them. Both skills are vital for building a culture of continuous improvement, but in the past collaboration may have been difficult because of the different approaches. This project, however, should help bring the two groups closer since it will alleviate some of the work load for QA/QC technicians and allow for greater collaboration.

In Stein, the project is being communicated as a project to automate Minitab analysis and consolidate data for the APR/PQR. It has been packaged as a project that will both drive productivity and help QA and production better understand their processes. The Global QA sees this project as an opportunity to entice the local sites to perform more detailed and frequent analysis of their data. There is an expectation that the time saved through automation will be

used to do more analysis and drive better process control. This is an important point and more information around it should be communicated in order to ensure there are no misunderstandings.

Finally, this tool will should enhance the existing culture of cooperation between local QA and production. The automated tool will save time and make it easier to quickly display trends. However, process improvement will only come about through continuous analysis of the data and determining route causes for excursions. This means production needs to be more involved with determining the features of this tool and its functionality. The culture in QA also clearly needs to shift from an organization that performs tests to one that can perform analysis and work more closely with production. This project is only the first step in that journey.

6.3 Political Lens

It is universally accepted within Novartis that automation and ease of use are paramount to the success of the project. Where politics can become an issue is in the choice of software engines for the project. IQP champions have invested time into training the staff at various sites in the use of Minitab which is specified in the current SOP. The IQP team has written the SOP for the accepted method of analysis and is therefore considered the authority on statistical process control for Novartis Pharmaceuticals. This provides the group with a lot of insight over the direction of the project at Stein, and their buy in with respect to the software platform is critical for the long term success of the tool.

At the local level the two key stakeholders are the QA department and production. Currently, all lab results are collected by the QA department and then presented to production. Deviations are first inspected to rule out lab error and if this is ruled out, a root cause investigation is opened. Production personnel then take over to determine the connection between the process and the data point. The automation of data consolidation and statistical analysis should allow production to see the data points at the same time as QA personnel. This might lead to production reacting without consulting QA. It is therefore important that the

current method for analyzing data and opening an investigation be reiterated to ensure that there is no confusion with the new tool.

At the corporate level, the main stakeholder is the Global QA department. They have tasked IT to help develop the automation of the APR/PQR but who still decide what the tools should look like and what capabilities it should have. The IQP department also plays an important advisory role in the statistical analysis portion of the project. While they don`t participate regularly in all project related meetings, their opinion is highly valued in negotiations about the direction of the project. At the same time it is important that steering committee members and developers be aware that the end users must be represented as well. We have asked lab technicians to provide suggestions to be passed on to both the global IT department and the global IQP group in order to ensure this occurs.

6.4 Conclusion

The three lens analysis above suggests that there are numerous opportunities to ensure long term success for the implementation of the project. The IQP group is a major stakeholder even though they may not be directly represented on the project team and their opinion must be factored into the strategy of the project. The research detailed in this thesis should help bring about change in both the statistical analysis procedure as well as in the choice of SPC software. Mindful of this, project leaders should approach the IQP group as well as the steering committee in advance of any recommendations so that these proposals can be screened for initial support before the steering committee convenes. It is also equally important to represent the views of lab technicians and process experts who will be using the tool. Project managers should ensure that local staff members have reasonable expectations about the level of automation in this new platform. For continuous monitoring, the graphs and deviations should be automated but in return this means QA will potentially have more responsibility for follow up. Another area that requires more attention is bringing local QA leadership and the project team leaders together and finalizing on the platform and user interface. It is important that senior leadership understand that the look and feel of the tool is as important as the analysis it performs. Along the same lines there needs to be a clear explanation of the technical restrictions with the different software

options that have been researched as part of this thesis. Since the steering committee operates at a very high level, it is important for the project team to present a comprehensive view of the software choice so that they fully appreciate the challenges with actually integrating an SPC engine into the automation project. To help with credibility on this front the project team must include senior technical experts who have experience in software and database projects. Finally, there should be a demonstration that will allow end users to see the software platform with new analytical techniques in place.

7 Conclusion and Next Steps

The IT platform for the Decision Support System together with the recommendations concerning statistical analysis will drastically reduce the time QA personnel spend on data consolidation and pave the way for better process control. Critical process parameters will be monitored on a daily basis (rather than annually) and by establishing internal specification at the six sigma limits (where appropriate), process exceptions will be flagged and investigated. QA personnel together with process experts will work together in investigating these exceptions and if a root cause is identified, the data point will be marked and removed from the analysis.

Continuous monitoring will also allow management to assess the process capabilities more frequently. In terms of the APR/PQR, the system should allow most of the report to be generated automatically. Process capability plots and run charts can be automatically generated and pasted into a MS Word document. Since the data will be monitored continuously, there should not be much time needed to investigate trends and modify the chart before the report is compiled. Data will be secured in the data warehouse and modification of this data will be restricted to ensure that the system meets auditing guidelines. The document management system will ensure that the report has a clear transparent audit trail with no loss of data. It will also ensure that the report remains in a secure database instead of versions of the report being transmitted via e-mail. The systems will also be scalable, which is important since this system may be deployed to multiple Novartis sites.

In terms of next steps, the IT system needs to be developed by the project management team and piloted at Stein or another Novartis Pharmaceutical manufacturing location. Staff should be trained to use the new system and a few products should be piloted for continuous monitoring. Dashboard templates should also be customized for the various end users who will be using the dashboard. One aspect of this project that must not be overlooked is the need for awareness and support. The success of this tool depends on users having access to all the features incorporated into the product as well as an efficient organizational framework to go with it. Management must decide how QA and production will collaborate to investigate exceptions and when internal OOE limits must be changed. There also needs to be clear communication between the two departments to ensure that process changes are communicated to the QA personnel who will be monitoring the data. With the right organizational incentives and the appropriate training,

the new system should ensure improved process performance as well as higher productivity for the QA department. It is also in-line with the FDA's desire for "A maximally efficient, agile, flexible pharmaceutical manufacturing sector that reliably produces high-quality drug products without extensive regulatory oversight" [32].

8 References

[1] Dean, D. and Bruttin, F. (2001). Productivity and the Economics of Regulatory Compliance in Pharmaceutical Production. Presented to FDA Science Board, Rockville, MD on Nov 16, 2001. Retrieved March 10th, 2008 from http://www.fda.gov/ohrms/dockets/ac/01/slides/3799s1_02_Dean.ppt.

[2] Scott, B. and Wilcock, A. (2008). Process Analytical Technology and Validation. *Validation of Pharmaceutical Processes, 3rd Addition,* Informa Healthcare USA.

[3] Winkle, Helen (2007, September 24). Implementing Quality by Design. Presented at PDA/FDA Joint Regulatory Conference. Retrieved on April 20th from http://www.fda.gov/Cder/OPS/ImplementingQualitybyDesign.pdf

[4] Femia, Dave. "Process Analytical Technology: Real-Time Reliability", Pharmaceutical Technology, November, 2005. Retrieved March 12, 2008 from: http://pharmtech.findpharma.com/pharmtech/Article/Process-Analytical-Technology-Real-Time-Reliabilit/ArticleStandard/Article/detail/261888.

[5] Novartis A G. (n.d.). company history. Retrieved March 10, 2010, from www.novartis.com: http://www.novartis.com/about-novartis/company-history/index.shtml

[6] Cook. J. "The 5 Steps of Starting PAT", Pharmaceutical Technology: Monitoring, Automation, and Control, 2007, 16-21. Retrieved March 3, 2008 from http://pharmtech.findpharma.com/pharmtech/article/articleDetail.jsp?id=407613&sk=&date=&pageID=2

[7] FDA (2009) Guidance for Industry Q10 Pharmaceutical Quality System April 2009 ICH http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073517.pdf

[7] ICH (1998) Q7: Good Manufacturing Practices for Pharmaceutical Ingredients Dated and endorsed by the Steering Committee on 5 February 1998

[8] Lee, Grazal "Product Annual/Quality Review: US–EU Comparative Analysis and Interpretations" *Pharmaceutical Technology* Vol. 3, No.32 http://pharmtech.findpharma.com/pharmtech/Peer-Reviewed+Research/Product-AnnualQuality-Review-USndashEU-Comparative/ArticleStandard/Article/detail/500406

[9] Yu, L. (2007). Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharmaceutical Research,* Vol. 25, No.4, pp.781-791.

[10] Buthmann, Arne. Dealing with Non-normal Data: Strategies and Tools. Retrieved November 2010. http://www.isixsigma.com/index.php?option=com_k2&view=item&id=1059:dealing-with-non-normal-data-strategies-and-tools&Itemid=208

[11] FDA. (1999). GUIDE TO INSPECTIONS OFQUALITY SYSTEMS. In FDA/Downloads. Retrieved December 2010, from http://www.fda.gov/downloads/ICECI/Inspections/UCM142981.pdf.

[12] Novartis Pharmaceutical SOP (2010). Process Capability Analysis using Minitab® Statistical Software.

[13] Montgomery, D.C., in: Introduction to Statistical Quality Control, Wiley, New York, 1993

[14] United States Pharmacopeia. General chapter 711: Dissolution. Rockville, Md.: United States Pharmacopeial Convention; 2009.

[15] United States Pharmacopeia. General chapter 1092: The dissolution procedure: development and validation. Rockville, Md.: The United States Pharmacopeial Convention; 2009.

[16] United States Pharmacopeia. General chapter 905: Uniformity of Dosage Units. Rockville, Md.: The United States Pharmacopeial Convention; 2009.

[17] United States Pharmacopeia,.Glossary of Terms Rockville, Md.: The United States Pharmacopeial Convention; 2009. http://www.usp.org/pdf/EN/USPNF/glossary.pdf

[18] The Free Pharmaceutical Encyclopedia, (2005). Stability Of Drugs:Factors Affecting Rates Of Hydrolysis And Oxidation. Retrieved November 2009 http://www.pharmpedia.com/Stability_Of_Drugs:Factors_Affecting_Rates_Of_Hydrolysis_And_Oxidation

[19] Yourstone, S., and W Zimmer (1992). Non-normality and the Design of Control Charts for Averages, *Decision Sciences*, Vol 32, No. 10, pp. 1099–1113.

[20] Duncan, A. J. (1978) "The Economic Design of p-charts to Maintain Current Control of a Process: Some Numerical Results," *Technometrics*, Vol 20, No. 1, pp. 235-243

[21] Montgomery, D.C (1982). "Economic Design of an x Control Chart" *International Journal of Quality & Reliability Management*, Vol. 14, No. 3, pp.234 - 259

[22] Borror, C. M., D. C. Montgomery, and G. C. Runger (1999). "Robustness of the EWMA Control Chart to Non-normality," *Journal of Quality Technology*, Vol. 31, No. 3, July 1999, pp. 309-316

[23] Burr, I. J. (1967). "The Effect of Nonnormality on Constants for x and R Charts," Industrial Quality Control, Vol 23. Issue May 1967, pp. 563-568

[24] Buthmann, A. (2000). ISIXSIGMA. In Making Data Normal Using Box-Cox Power Transformation. Retrieved November 2010.

[25] Farnum, N (1996-97). "Using Johnson Curves to Describe Non-Normal Data", *Quality Engineering*, Vol. 9, No.2, pp. 329- 336,

[26] Chou, Y, Polansky, A and Mason, R.L (1998). "Transforming Nonnormal Data to Normality in Statistical Process Control," *Journal of Quality Technology*, Vol. 30, No. 2, pp. 133-141

[27] Clements, J. A. (1989). "Process Capability Calculations for Non-Normal Distributions" *Quality Progress*, Issue September 1989, pp.95-100

[28] Willemain, T.R. and Runger G. C. (1996). "Designing Control Charts Based on Empirical Reference Distribution" Vol. 28, No. 1, January 1996, pp. 31-38

[29] Walton, S. Cline, A. (2000). Citing Website. In *Data Warehouses/Data Marts Repositories for Data Mining*. Retrieved March 2011, from http://www.carolla.com/wp-dw.htm.

[30] Johnson, L. (2007) "Modeling Non-normal Data Using Statistical Software" *R&D Magazine,* August 2007

[31] Bower, K (1998), "Some Misconceptions about the Normal Distribution" American Society for Quality, Six Sigma Forum, May 2003

[32] Janet Wookcock (Director CDER), 2008 CMC Conference.

[33] Luh, W. (2002). "Using Johnson's Transformation with Approximate Test Statistics for the Simple Regression Slope Homogeneity" *Journal of Experimental Education*, Vol.71, No.1 pp 69-81